ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

Learning from Partially Labeled Data

Siamak Mehrkanoon[*,1], Xiaolin Huang[2], Johan A.K. Suykens[3]

1- Department of Data Science and Knowledge Engineering,
Maastricht University, The Netherlands.
2- Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University
3- Department of Electrical Engineering ESAT-STADIUS,
KU Leuven, B-3001 Leuven, Belgium
[*]Corresponding author, e-mail: siamak.mehrkanoon@esat.kuleuven.be

Abstract.
Providing sufficient labeled training data in many application domains is a laborious and costly task. Designing models that can learn from partially labeled data, or leveraging labeled data in one domain and unlabeled data in a different but related domain is of great interest in many applications. In particular, in this context one can refer to semi-supervised modelling, transfer learning, domain adaptation and multi-view learning among others. There are several possibilities for designing such models ranging from shallow to deep models. These type of models have received increasing interest due to their successful applications in real-life problems. This paper provides a brief overview of recent techniques in learning from partially labeled data.

1 Introduction

In many application domains one encounters limited labeled training data while unlabeled data are more easily generated and available. The literature has witnessed several attempts in overcoming the challenges of learning from partially labeled datasets by proposing models that can leverage available labeled or unlabeled data in different domains. The semi-supervised learning, domain adaption, multi-view learning are among existing proposed models.

Semi-supervised models use both labeled and unlabeled data points in the learning process. To this end, one incorporates the labels/unlabeled data in the learning process to enhance the clustering/classification performance. Usually both labeled and unlabeled data are from one domain and are drawn from the same distribution. However there are several cases in which they are drawn from different distribution, they have different feature dimensions. It is also possible that they are drawn from two different but related domains. One can further generalize this pattern and consider the case that the training and test data do not exhibit the same distribution, same feature domains or their statistical properties change over time. The data might also have been described using different representations, views or modalities. In all these cases, many classical machine learning algorithms fail to provide the desired classification performance. Therefore, addressing these challenging problems by designing suitable techniques and models has recently attracted many researchers. In what follows we will give an overview of some of the recent existing techniques that are used for learning from

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

partially labeled data with above mentioned level of complexity. In particular, in Section 2, a brief overview of existing semi-supervised modelling techniques is provided. Domain adaptation methodologies are discussed in Section 3. Muti-view learning methodologies are given in section 4.

2 Semi-supervised learning

Traditionally, there are two different types of tasks in machine learning, namely supervised and unsupervised. In supervised learning all the training samples are labeled and one tries to learn a right mapping from the given input to the desired output. On the other hand, in unsupervised learning the training samples are all unlabeled and the task is to for instance group or cluster the given data and find the underlying patterns of the data. Semi-supervised learning (SSL) lies between supervised and unsupervised learning where it aims at learning from both labeled and unlabeled data, see Fig. 1. Learning from few labeled and a large number of unlabeled training data is directly relevant to several practical problems where it is expensive to produce labeled data. Semi-supervised learning (SSL) is one of the modelling approaches that is able to cope with the limited training data and has already shown its importance in many application domains including automatic classification of web pages, medical images analysis, traffic or climate data, biomedical signals among others. Int this type of learning one can benefit from different type of partial supervision.

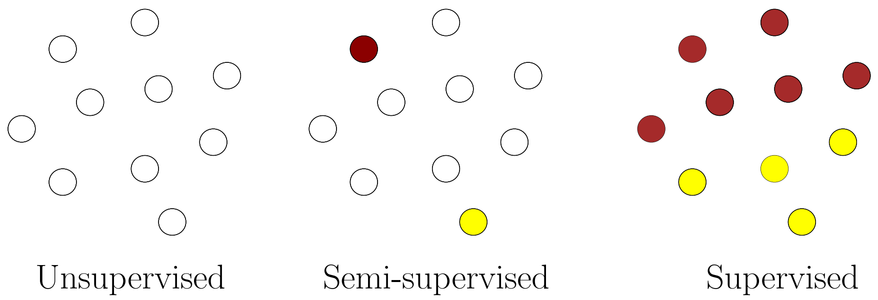

Fig. 1: Supervision Spectrum

For instance, there might be constraints on the data points such as having the same target or not, or simply the target values for only some of the training samples are given. In general there are two views on semi-supervised modeling. In the first view the SSL is seen as unsupervised learning guided by constraints and in the second view, SSL is considered as supervised learning with additional information on the distribution of the examples [1, 2, 3, 4, 5, 6]. The semi-supervised learning can be classified into two categories, i.e. transductive and inductive learning [7]. In transductive learning the labels for a specified set of test data are predicted. In inductive learning one learns a decision function from a training data for future unobserved test instances. One can further categorize the semi-supervised inductive learning into semi-supervised clustering and classification [2]. In semi-supervised clustering the task is to exploit labeled data to adjust the cluster memberships of the unlabeled data. In contrary, semi-supervised classification uses both unlabeled and labeled data to obtain a better predictions and classification accuracy on unseen test data points [2].

Traditionally, in semi-supervised techniques, a classifier is first trained using the available labeled data points and then the classified unlabeled data with the highest confidence scores are added incrementally to the training set and the process is repeated until the convergence criteria is met [8]. Some of the successful semi-supervised models that have been proposed in the literature are for instance, the Laplacian support vector machine (LapSVM) [7] with manifold regularization, local spline regression for semi-supervised classification [9], A label propagation [10], nonlinear embedding in deep multi-layer architectures [11] and Semi-Supervised Kernel spectral clustering [2]. The design of many existing SSL models benefits form one of the cluster assumption or manifold assumption. In the former, the decision boundary between classes should lie in the lower density region of the space. The latter assumes that the target function should change smoothly along the tangent direction. It should be noted that most of the graph based models perform transductive learning. However, LapSVM is one of them with a data-dependent geometric regularization which has the out-of-sample extension property and therefore is able to perform inductive inference.

The authors in [2] formulated a regularized Kernel Spectral Clustering (KSC) which can be operated for both semi-supervised classification and clustering. The KSC which is a completely unsupervised algorithm is used as the core model and the available labeled data is incorporated to this model via adding a regularization term to the cost function of the KSC formulation. Thanks to the added regularization term, the model clustering scores are adjusted to be as close as possible to the desired labels. The model is then obtained by solving a linear system of equations in the dual. Extension of this model to analyze streaming data in an online fashion has been proposed in [12].

Nowadays, deep learning has shown its great power in several learning tasks. When training deep neural networks, samples with qualified labels are quite important, due to their over-fitting essence. However, there are many applications which contain vast amount of unlabelled data. For example, in computer assisted diagnosis tasks, there could be many medical images, but the labels, especially accurate masks, need very heavy human consumption. Thus, learning from partially labeled data becomes attractive for training deep neural networks for those tasks. There are different ways to use unlabeled data for training neural networks. The unlabeled data could pre-train parts of the neural networks to coincide with some expectations, e.g., sparse structure or similarity in latent space. In those methods, unlabeled data are used in a unsupervised manner [13, 14]. There are various methodologies for giving labels to unlabeled data. In [15], a surrogate class is generated by automatically generating samples from unlabeled data. In [16, 17, 18, 19], pseudo-labels are given to the unlabeled data by links between labeled and unlabeled samples, which could be represented by

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

graph or manifold.

In addition to predicting and assigning labels to unlabeled data and transferring the training task to supervised one, designing loss/target functions for those unlabeled data is another efficient way. The design could be rooted in underlying guesses, observations, or reasonable assumptions. For example, the output distribution of a neural network for labeled and unlabeled data should be the same [20]; the entropy should be minimized not only on labeled data but also on the unlabeled ones [21]; all the data, no matter there are labels or not, could be embedded in the same manifold [22]; a generative model could be trained on unlabeled data and be mixed up with labelled data [23]. Among many deep semi-supervised methods, the ladder net, proposed by [24], is a representative work and has achieved very good performance. The idea of ladder net appeared first in unsupervised learning [25] and the basic structure is an auto-encoder, which is to remove noise by minimizing the difference of the original input and its reconstruction. In semi-supervised learning, the label from the labeled and the structure in all data are used together in the ladder net such that clean features coinciding with prior expectations could be learned [24].

3 Domain adaptation

In many practical machine learning problems the statistical properties of the data change from one domain to another. This domain shifts brings new challenges for classical machine learning algorithms which are designed based on the assumption stationary data. Therefore the need for designing models that can leverage the information of the labeled data in one domain to better classify unseen data in a target domain is highly desirable. To this end, domain adaptation based models are proposed in the literature. In both domain adaption and semi-supervised learning one tries to leverage labeled data to improve the model generalization performance on the unseen unlabeled data. However, in domain adaptation there are extra assumptions, i.e. one is dealing with two domains which exhibit different distributions and/or there is a feature dimension mismatch. In case one can ignore these assumptions, then the source and target domain can provide the labeled and unlabeled data to be used by semi-supervised models.

There are three main scenarios one can consider for domain adaption problems which differ in the information considered for the target task. In unsupervised domain adaptation, one learns from a set containing labeled, unlabeled source instances and unlabeled target instances, i.e the labeled information from the target domain is not taken into account. In supervised domain adaptation, all the instances considered are labeled. In the semi-supervised setting, one learns from labeled source instances as well as a small fraction of the target labeled instances [26]. The main two types of domain adaptation problems that are addressed in the literature are homogeneous and heterogeneous domain adaptation. In the homogeneous domain adaptation, the feature representation for the source and target domains is the same, i.e. $\mathcal{X}_s = \mathcal{X}_t$ but the marginal probabil-

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

ity distributions are not the same i.e. $P(\mathcal{X}_s) \neq P(\mathcal{X}_t)$. However heterogeneous domain adaptation is more challenging since the distributions, feature domains and dimensions in source and target are different. The literature have witnessed several domain adaptation methodologies for addressing both homogeneous and heterogeneous data. The two main categorize are models with shallow and deep architectures. In what follows we give a brief overview of some of the successful shallow domain adaptation methods.

In shallow and traditional models there are three major directions of works proposed in the literature for domain adaptation problems including instance reweighting methods [27, 28, 29] and feature transformation [30, 31]. In the sample reweighting approach, one assigns sample-dependent weights to the training data with the aim of minimizing the distribution discrepancy between the source and target data points in the re-weighted space [27]. In practice one usually can use the density ratio between the probability densities of the two domains to estimate the sample dependent weights. Another main approach in domain adaptation is to learn a domain-invariant feature representation for both source and target domains. Several successful feature transformation methods are proposed in the literature. The authors in [30] introduced the Transfer Component Analysis (TCA) in order to learn common latent features having the same marginal distribution across the source and target domains. The Structural Correspondence Learning method proposed in [31] learns a common feature space by identifying correspondence among features from different domains. A domain adaptation approach that uses the correlation subspace as a joint representation of the source and target data is introduced in [32].

Often in many domain adaptation problems, side information in the form of correspondence (paired) instances are available across domains, see Fig. 2. In this case, the instance similarity constraints between domains, can be used to enhance the generalization performance of the model [33]. A Regularized Semi-Paired Kernel Canonical Correlation Analysis (RSP-KCCA) model is introduced in [26] for learning a new representation of the data for the sake of the domain adaptation problem. This model also belongs to the family of feature transformation methods. The corresponding optimization problem is formulated in the primal-dual Least Squares Support Vector Machines setting where side information are incorporated through proper regularization terms. A new representation of both source and target data sets is learnt by solving a linear system of equations in the dual. The model seamlessly integrates unlabeled, labeled, paired and unpaired instances and can be employed in unsupervised, semi-supervised as well as supervised fashions.

Other types of domain adaptation models that have recently gained attentions poses deeper architectures and are based on artificial neural-networks. Within this category of models one can refer to learning representations using the marginalized stacked denoising autoencoders (mSDA) [34], adjusting the the pre-trained networks to the new task by fine-tuning [35, 36, 37], interpolating between domains [38] and cross-domain neural-kernel networks [39].

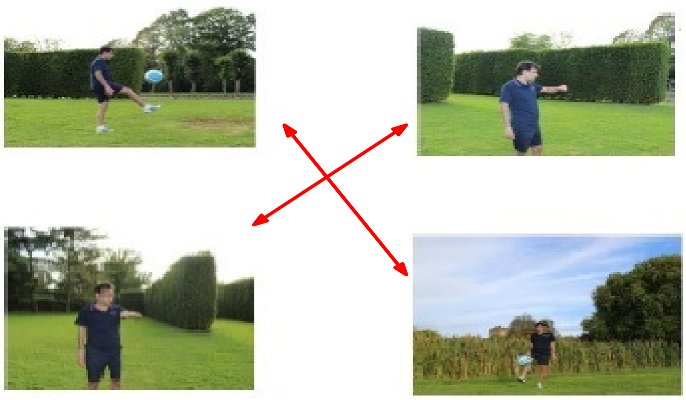

Fig. 2: Example of labeled paired instances [26].

4 Multi-view learning

The success of learning from partially labelled data relies on the full use of the links between labelled and unlabelled data. One interesting and effective way is multi-view learning, which was first proposed by [40]. Its basic assumption is that labeled and unlabeled data share similarity in different views. The original idea of multi-view learning, called co-training [40], is to learn several weak classifiers based on labeled data but from multiple views and then bootstrap the weak classifiers by unlabeled data. Later on, the idea of co-training has been theoretically analyzed in [41, 42] and successfully applied in other tasks, e.g., spectral clustering [43] [44]. Another way of using multi-view data is to design a regularization term that could capture the joint prior knowledge. In this direction, [45] is a representative method that applies kernel canonical correlation analysis [46] in support vector machines to model the similarity in a common latent space. Establishing a co-regularization term for multiple views is applicable and effective for many tasks. For example, [46] designed a regularization term from manifold-based ensemble learning; the multi-view version of the Hessian regularization method [47] achieves promising performance in image annotation; the application on graph-oriented method could be found in [48, 49, 50].

Here we list multi-view learning as a category of learning from partially labeled data. But this idea is actually applicable for many applications, no matter in supervised, semi-supervised, or unsupervised tasks. For example, the idea of co-regularization can used in unsupervised learning task [51] when there exists a good common base manifold. In the person re-identification problem, different cameras provide multi-view for the same person and multi-view learning that considers different views simultaneously is helpful [52, 53, 54]

5 Conclusions

We have briefly reviewed several existing approaches in connection to learning from partially labeled data. The covered approaches range from shallow models to deep architectures in the context of semi-supervised learning, domain adaptation and multi-view learning.

Acknowledgments
This work was supported by the Postdoctoral Fellowship of the Research Foundation-Flanders (FWO: 12Z1318N). National Natural Science Foundation of China (No. 61977046). EU: The research leading to these results has received funding from the European Research Council under the European Unionâs Horizon 2020 research and innovation program / ERC Advanced Grant E-DUALITY (787960). This paper reflects only the authorsâ views and the Union is not liable for any use that may be made of the contained information.Research Council KUL: Optimization frameworks for deep kernel machines C14/18/068. Flemish Government: FWO: projects: GOA4917N (Deep Restricted Kernel Machines: Methods and Foundations), PhD/Postdoc grant. Impulsfonds AI: VR 2019 2203 DOC.0318/1QUATERKenniscentrum Data en Maatschappij. Ford KU Leuven Research Alliance Project KUL0076. Siamak Mehrkanoon is an assistant professor at Maastricht University. Xiaolin Huang is an associate professor at Shanghai Jiao Tong University. Johan Suykens is a full professor at the KU Leuven, Belgium.

References

[1] Siamak Mehrkanoon and Johan A K Suykens. Non-parallel semi-supervised classification based on kernel spectral clustering. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 2311–2318. IEEE, 2012.

[2] Siamak Mehrkanoon, Carlos Alzate, Raghvendra Mall, Rocco Langone, and Johan A K Suykens. Multiclass semisupervised learning based upon kernel spectral clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 26(4):720–733, 2015.

[3] Gustavo Camps-Valls, Tatyana V Bandos Marsheva, and Dengyong Zhou. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3044–3054, 2007.

[4] Siamak Mehrkanoon and Johan AK Suykens. Multi-label semi-supervised learning using regularized kernel spectral clustering. In *In Proc. of IEEE World Congress on Computational Intelligence (WCCI-IJCNN 2016)*, pages 1233–1240, 2016.

[5] Siamak Mehrkanoon and Johan A K Suykens. Large scale semi-supervised learning using KSC based model. In *Proc. of International Joint Conference onNeural Networks (IJCNN)*, pages 4152–4159, 2014.

[6] Yi Liu, Rong Jin, and Liu Yang. Semi-supervised multi-label learning by constrained non-negative matrix factorization. In *Proceedings of the national conference on artificial intelligence*, volume 21, page 421. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[7] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research*, 7:2399–2434, 2006.

[8] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. *Semi-Supervised Learning*. MIT Press, 2010.

[9] Shiming Xiang, Feiping Nie, and Changshui Zhang. Semi-supervised classification via local spline regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2039–2053, 2010.

[10] M. Karasuyama and H. Mamitsuka. Multiple graph label propagation by sparse integration. *IEEE Transactions on Neural Networks and Learning Systems*, 24(12):1999–2012, 2013.

[11] Jason Weston, Frédéric Ratle, Hossein Mobahi, and Ronan Collobert. Deep learning via semi-supervised embedding. In *Neural networks: Tricks of the trade*, pages 639–655. Springer, 2012.

[12] Siamak Mehrkanoon, Oscar Mauricio Agudelo, and Johan A K Suykens. Incremental multi-class semi-supervised clustering regularized by kalman filtering. *Neural Networks*, 71:88–104, 2015.

[13] Koray Kavukcuoglu, Marc'Aurelio Ranzato, and Yann LeCun. Fast inference in sparse coding algorithms with applications to object recognition. *arXiv preprint arXiv:1010.3467*, 2010.

[14] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1422–1430, 2015.

[15] Alexey Dosovitskiy, Jost Tobias Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with convolutional neural networks. In *Advances in neural information processing systems*, pages 766–774, 2014.

[16] Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 2, 2013.

[17] Weiwei Shi, Yihong Gong, Chris Ding, Zhiheng MaXiaoyu Tao, and Nanning Zheng. Transductive semi-supervised deep learning using min-max features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 299–315, 2018.

[18] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5070–5079, 2019.

[19] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. S4l: Self-supervised semi-supervised learning. In *Proceedings of the IEEE international conference on computer vision*, pages 1476–1485, 2019.

[20] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems*, pages 1195–1204, 2017.

[21] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993, 2018.

[22] Philip Haeusser, Alexander Mordvintsev, and Daniel Cremers. Learning by association–a versatile semi-supervised training method for neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 89–98, 2017.

[23] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *Advances in Neural Information Processing Systems*, pages 5050–5060, 2019.

[24] Antti Rasmus, Mathias Berglund, Mikko Honkala, Harri Valpola, and Tapani Raiko. Semi-supervised learning with ladder networks. In *Advances in neural information processing systems*, pages 3546–3554, 2015.

[25] Harri Valpola. From neural pca to deep unsupervised learning. In *Advances in independent component analysis and learning machines*, pages 143–171. Elsevier, 2015.

[26] Siamak Mehrkanoon and Johan AK Suykens. Regularized semipaired kernel CCA for domain adaptation. *IEEE transactions on neural networks and learning systems*, 2017.

[27] Masashi Sugiyama, Shinichi Nakajima, Hisashi Kashima, Paul V Buenau, and Motoaki Kawanabe. Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in neural information processing systems*, pages 1433–1440, 2008.

[28] Wen-Sheng Chu, Fernando De la Torre, and Jeffery F Cohn. Selective transfer machine for personalized facial action unit detection. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3515–3522. IEEE, 2013.

[29] Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90(2):227–244, 2000.

[30] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2011.

[31] John Blitzer, Ryan McDonald, and Fernando Pereira. Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 120–128. Association for Computational Linguistics, 2006.

[32] Yi-Ren Yeh, Chun-Hao Huang, and Yu-Chiang Frank Wang. Heterogeneous domain adaptation and classification by exploiting the correlation subspace. *IEEE Transactions on Image Processing*, 23(5):2009–2018, 2014.

[33] Jeff Donahue, Judy Hoffman, Erik Rodner, Kate Saenko, and Trevor Darrell. Semi-supervised domain adaptation with instance constraints. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 668–675. IEEE, 2013.

[34] Minmin Chen, Zhixiang Xu, Kilian Weinberger, and Fei Sha. Marginalized denoising autoencoders for domain adaptation. In *Proceedings of the 29th International Conference on International Conference on Machine Learning*, pages 1627–1634. ICML, 2012.

[35] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[36] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1717–1724. IEEE, 2014.

[37] Brian Chu, Vashisht Madhavan, Oscar Beijbom, Judy Hoffman, and Trevor Darrell. Best practices for fine-tuning visual classifiers to new domains. In *European Conference on Computer Vision*, pages 435–442. Springer, 2016.

[38] Sumit Chopra, Suhrid Balakrishnan, and Raghuraman Gopalan. Dlid: Deep learning for domain adaptation by interpolating between domains. In *ICML Workshop on Challenges in Representation Learning.*

[39] Siamak Mehrkanoon. Cross-domain neural-kernel networks. *Pattern Recognition Letters*, 125:474–480, 2019.

[40] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100, 1998.

[41] Kamal Nigam and Rayid Ghani. Analyzing the effectiveness and applicability of co-training. In *Proceedings of the ninth international conference on Information and knowledge management*, pages 86–93, 2000.

[42] Avrim Blum and Yishay Mansour. Efficient co-training of linear separators under weak dependence. In *Conference on Learning Theory (COLT)*, volume 65, 2017.

[43] Lynn Houthuys, Rocco Langone, and Johan AK Suykens. Multi-view kernel spectral clustering. *Information Fusion*, 44:46–56, 2018.

[44] Ulf Brefeld and Tobias Scheffer. Semi-supervised learning for structured output variables. In *Proceedings of the 23rd international conference on Machine learning*, pages 145–152, 2006.

[45] Jason Farquhar, David Hardoon, Hongying Meng, John S Shawe-Taylor, and Sandor Szedmak. Two view learning: Svm-2k, theory and practice. In *Advances in neural information processing systems*, pages 355–362, 2006.

[46] David R Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.

[47] Weifeng Liu and Dacheng Tao. Multiview hessian regularization for image annotation. *IEEE Transactions on Image Processing*, 22(7):2676–2687, 2013.

[48] Feiping Nie, Guohao Cai, and Xuelong Li. Multi-view clustering and semi-supervised classification with adaptive neighbours. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[49] Meng Wang, Weijie Fu, Shijie Hao, Hengchang Liu, and Xindong Wu. Learning on big graph: Label inference and regularization with anchor hierarchy. *IEEE transactions on knowledge and data engineering*, 29(5):1101–1114, 2017.

[50] Yucen Luo, Jun Zhu, Mengxi Li, Yong Ren, and Bo Zhang. Smooth neighbors on teacher graphs for semi-supervised learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8896–8905, 2018.

[51] Yifeng Fan, Tingran Gao, and Zhizhen Jane Zhao. Unsupervised co-learning on g-manifolds across irreducible representations. In *Advances in Neural Information Processing Systems*, pages 9038–9050, 2019.

[52] Sheng Li, Ming Shao, and Yun Fu. Person re-identification by cross-view multi-level dictionary learning. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2963–2977, 2017.

[53] Dapeng Tao, Yanan Guo, Baosheng Yu, Jianxin Pang, and Zhengtao Yu. Deep multi-view feature learning for person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10):2657–2666, 2017.

[54] Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng. Cross-view asymmetric metric learning for unsupervised person re-identification. In *Proceedings of the IEEE international conference on computer vision*, pages 994–1002, 2017.